

# Introducing the four dimensions 4D migraine scale: A composite score proposal evaluating migraine severity and treatment efficacy

Cephalalgia

2025, Vol. 45(12) 1–12

© International Headache Society 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03331024251398020

journals.sagepub.com/home/cep



Patrizio Pasqualetti<sup>1</sup> , Claudia Altamura<sup>2,3</sup> , Luisa Fofi<sup>2</sup> ,  
Luigi Francesco Iannone<sup>4,5</sup> , Marilena Marcosano<sup>2</sup> ,  
Giulia Bellini<sup>3</sup>, Marina Romozzi<sup>6,7</sup> , Sabina Cevoli<sup>8</sup>,  
Gian Camillo Manzoni<sup>9</sup> and Fabrizio Vernieri<sup>2,3</sup> 

## Abstract

**Background:** Different parameters are currently used to evaluate migraine frequency and disability. We aimed to formulate a composite scale including the most relevant clinical measures to better evaluate the burden of migraine.

**Methods:** To create the composite four dimensions 4D migraine scale, we selected the most commonly used outcome measures: monthly migraine days (MMDs), number of monthly acute medications (MAMs), pain intensity (by Numerical Rating Score, NRS) and Migraine Disability Assessment (MIDAS) Score. Each parameter was categorized in different levels: five for MMDs, seven for MAMs, five for NRS and six for MIDAS to cover the entire empirical range of each variable. First, the relative weight of each level per parameter was rated by 197 migraine patients and 118 headache experts using Conjoint Analysis. Secondly, we applied the 4D migraine score to a sample of patients treated with galcanezumab. We assessed its concurrent validity for the scale's single parameters and the Head Impact Test HIT-6, an external patient-reported outcome measure.

**Results:** There was a substantial agreement between clinicians and patients about the weight of each parameter in terms of Relative Importance (RI). For both categories, MMDs were the most relevant attribute (RI: 34% for clinicians, 32% for patients) and pain intensity NRS the least important (RI: 14% vs 13%). Though marginally, MIDAS was more important than MAMs for patients (29% vs. 26%), while for clinicians the relevance of these two attributes was almost equal (26% and 27%). In terms of the utility assigned to each level, strong agreement was confirmed between clinicians and patients. According to the utilities implicitly attributed by participants to the chosen representative levels of the four parameters, four different statistical models were derived, allowing to compute utilities from all possible values of MMDs, MAMs, NRS and MIDAS and finally a unique 4D migraine score for every possible patient, ranging from 0 (without migraine) to 100 (with the most severe migraine). The 4D score was valid in terms of sensitivity to changes and showed concurrent validity with respect to HIT-6.

<sup>1</sup>Section of Health Statistics and Biometry, Department of Public Health and Infectious Diseases, Sapienza University of Rome, Roma, Italy

<sup>2</sup>Headache and Neurosonology Unit, Fondazione Policlinico Universitario Campus Bio-Medico, Roma, Italy

<sup>3</sup>Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Roma, Italy

<sup>4</sup>Digital and Predictive Medicine, Pharmacology and Clinical Metabolic Toxicology-Headache Center and Drug Abuse-Laboratory of Clinical Pharmacology and Pharmacogenomics, AOU Policlinico Di Modena, Modena, Italy

<sup>5</sup>Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

<sup>6</sup>Dipartimento Universitario di Neuroscienze, Università Cattolica del Sacro Cuore, Roma, Italy

<sup>7</sup>Neurologia, Dipartimento di Neuroscienze, Organi di Senso e Torace, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Roma, Italy

<sup>8</sup>Programma Cefalee e Algie facciali, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy

<sup>9</sup>Poliambulatorio Dalla Rosa Prati, Centro Diagnostico Europeo, Garofalo Health Care, Parma, Italy

## Corresponding author:

Fabrizio Vernieri, Headache and Neurosonology Unit, Fondazione Policlinico Campus Bio-Medico, via Alvaro del Portillo 200, 00128 Roma, Italy.

Email: f.vernieri@policlinicocampus.it



**Conclusion:** The 4D migraine scale, based on the preference weights of both clinicians and patients, could be useful to fully quantify the migraine burden and the efficacy of a treatment.

### Keywords

migraine, discrete choice experiment, composite score, patient preference, preventive treatment

Date received: 26 June 2025; revised: 21 October 2025; accepted: 26 October 2025

## Introduction

Different outcome measures are used nowadays to evaluate the severity of migraine. Among them, migraine frequency is the most commonly employed, as generally the higher the number of monthly migraine days (MMDs), the higher the burden. Migraine frequency is also usually used for diagnostic criteria, subdividing episodic migraine (EM), i.e., up to 14 MMDs, and chronic migraine (CM), i.e., 15 or more monthly headache days (MHDs) (1). However, the CM criteria are the only ones reported in the International Classification of Headache Disorders (2) (1.3 ICHD-3, 2018), where attack frequency is specified (i.e., at least eight days/month must have the features of migraine headache). Nevertheless, the use of a 15-day/month threshold to distinguish EM and CM has been criticized as it seems neither to fully capture the burden of illness nor to reflect the treatment needs of patients (3,4). Almost all studies consider MMDs as the primary endpoint to evaluate patients' migraine burden and pharmacological treatment efficacy (5,6). Many studies, however, also took into consideration MHDs, especially in people with CM, as it is almost impossible to distinguish the individual episodes of headache in patients with daily or near daily headaches, and the characteristics of the headache may change not only from day to day but even within the same day (2).

Clinicians may also assess migraine burden by the number of days on which patients take at least one acute medication, or alternatively by the total number of symptomatic drugs patients took during the observation period. Indeed, patients with migraine may take multiple acute medications on the same day to treat severe and long-lasting attacks; on the other hand, patients with mild and short-lasting headaches might not take any acute medication on some headache days. Single-attack pain intensity, or rather the mean intensity of attacks, and different evaluation scales were often considered in evaluating the degree of migraine pain (7). Thus, the use of a headache diary is fundamental for reporting days, drugs, and pain intensity of the attacks to be discussed between patients and physicians during clinical examinations.

However, the burden of migraine cannot be fully assessed through measures focusing exclusively on headache frequency and pain degree or acute medication intake. The

concept of headache-related functional disability and impact on quality of life (QoL) reported by patients has become of paramount importance in fully defining the migraine burden. Migraine often has a dramatic impact on people's ability in different settings: i.e., work or school, housework, family, and leisure activities. The higher the frequency or intensity of headache attacks, the greater the impairment on functional ability and quality of life. Information about these aspects of migraine helps physicians to fully evaluate the need for treatment. Patient-reported outcome measures (PROMs) assessing headache-related disability and quality of life are now recognized as essential tools in evaluating the impact of migraine in clinical practice (8). Different validated questionnaires or scales are available to measure disability and the impact on QoL of migraine (9–11). However, studies often use different PROMs or those are not even considered, even in randomized controlled trials or real-life studies.

Nevertheless, either reporting frequency, pain intensity, or drug intake, or assessing disability or quality of life, could not entirely represent the burden of migraine. It would be important to have a composite measure, including the most relevant parameters from those reported above, and designed employing a weighted approach to fully evaluate the migraine severity and the degree of treatment efficacy. Different attempts have been made, also recently (12,13), but they considered only migraine frequency or attack features (e.g., total pain burden, etc.), failing to also include disability and quality of life.

The present study aimed to propose a composite scale including the most common parameters used to evaluate migraine frequency, pain intensity, acute medication intake, and patient-reported disability, and to apply the obtained score for evaluating treatment response in patients on a monoclonal antibody anti-CGRP (galcanezumab), and assessing its concurrent validity compared to the single parameters.

## Methods

The composite scale includes four of the most common parameters used to evaluate migraine severity: frequency,

pain intensity, acute medication intake, and a patient-reported disability measure.

The relative weight of each parameter was rated based on both patients' and headache experts' opinions, who responded to an anonymous web-based Discrete Choice Experiment (DCE) survey using a validated approach.

Therefore, we applied the obtained composite score to assess its suitability in monitoring treatment response in a post-hoc analysis of the prospective observational GARLIT study (14) conducted on a sample of EM and CM patients treated with a monoclonal antibody anti-CGRP (galcanezumab). We assessed its concurrent validity with respect to the four parameters considered singularly and a different measure (Head Impact Test HIT-6 Scale) (10) not included in the composite scale.

### Participants

Adult migraine patients under the care of the Headache Center of the Fondazione Policlinico Campus Bio-Medico and/or members of Alleanza Cefalalgici (Al.Ce.), an association of migraine patients in Italy, were invited to fill the web-based Discrete Choice Experiment survey.

The clinicians taking the survey were headache experts working in several Italian headache centers and affiliated to either the two Headache Societies in Italy, Associazione Neurologica Italiana Ricerca sulle Cefalee ANIRCEF and Società Italiana per lo Studio delle Cefalee (SISC).

Enrollment in the study was initiated when individuals accessed the web-based survey link. Upon accessing the survey, participants were provided with an information page outlining the study's purpose and the anonymous nature. Consent to participate was obtained electronically by ticking a checkbox before continuing with the survey.

Enrollment and inclusion/exclusion criteria of the GARLIT study participants are described elsewhere (14).

**Discrete choice experiment.** We chose four of the best known parameters commonly used as migraine outcome measures: Monthly Migraine Days (MMDs) for measuring Days with migraine, Monthly Acute Medications (MAMs) for counting number of symptomatic Drugs taken, the worst attack pain intensity (by Numerical Rating Score, NRS, score 0–10) for evaluating pain Degree (or Dolor in Latin) and Migraine Disability Assessment (MIDAS) questionnaire for scoring Disability, to create the composite four dimensions 4D migraine scale (where 4D stands for Days, Drugs, pain Degree and Disability).

Each parameter was categorized into a number of levels: five for MMDs (1,7,14,21,28), seven for MAMs (1,7,14,21,28,35,42), five for NRS (1,3,5,7,9), and six for MIDAS (3,8,16,25,50,90). The selection of these levels was dictated by the need to cover the entire empirical range of each scale, considering that two of them have predefined bounds (0–28 for MMDs, 0–10 for NRS), one (MAMs) has no predefined upper limit (for instance a patient with the

maximal value of MMD=28 days which assumes at least two drugs for each migraine day would result in a MAMs=56, i.e.,  $\geq 40$ ), thus we chose two levels beyond 28 (35 and 42), and, finally, one (MIDAS) has a theoretical upper bound (270) but usually its distribution – before taking an effective therapy – is typically always highly skewed and a score of 90 is usually considered indicative of a very strong disability due to migraine. Therefore, the levels were chosen by experienced migraine clinicians to indicate, for each parameter, the representative degrees of severity, taking into consideration real-world data from 171 patients at T0 (before treatment), where MAMs and MIDAS histograms are represented after log-transformation to account for their log-normal distribution (Figure 1).

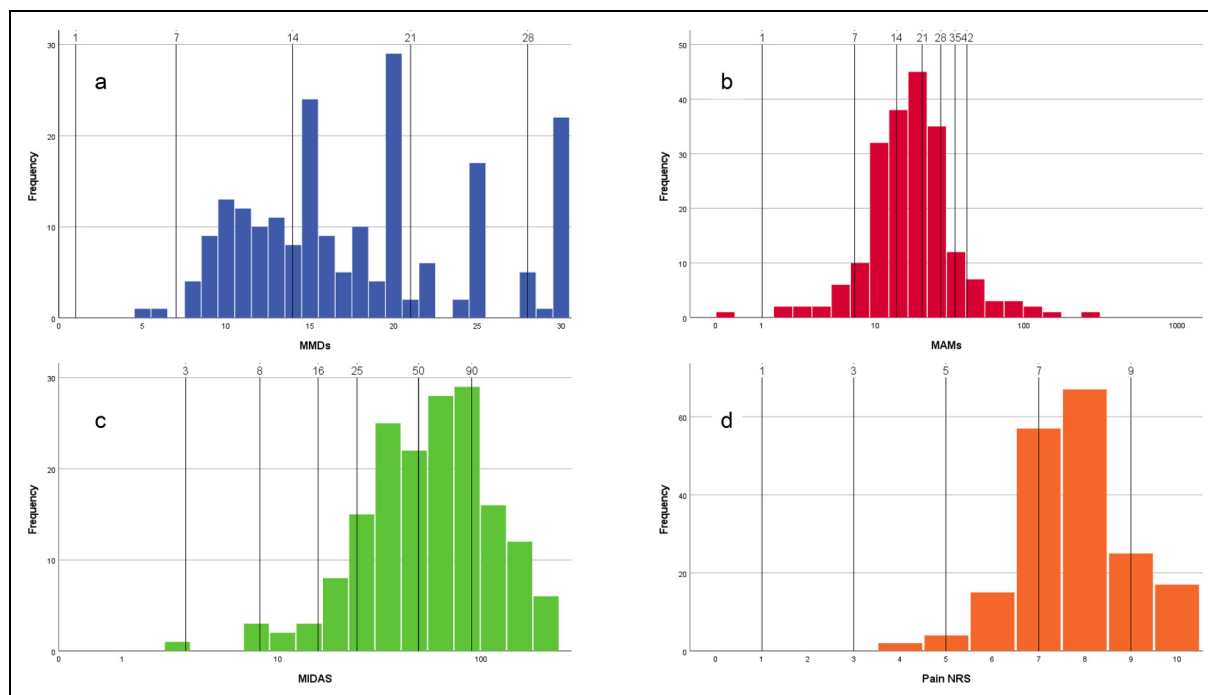
The chosen parameters – referred to as “attributes”, according to the Discrete Choice Experiments (DCE) terminology (15) – and levels within each attribute are reported in Table 1 (DCE design).

Using Lighthouse Studio (<https://sawtoothsoftware.com/>), an advanced platform for Conjoint Analysis, a series of pairwise choice tasks was generated. For each of them, two hypothetical (but realistic) patients suffering from migraine were described utilizing scale parameters. In each task, respondents (patients and clinicians) were asked to assess which of the patients was in better health (Figure 2).

Each profile is randomly generated to obtain a representative subset of all possible profiles and, consequently, of the possible pairwise comparisons. The optimal DCE design should be orthogonal (if all attribute levels vary independently and thus are not correlated) and balanced (when each level of an attribute occurs the same number of times). In this study, a perfectly orthogonal and balanced design was not feasible, as it would have generated several unrealistic profiles. Examples of implausible combinations are: a patient suffering one day with migraine a month (MMD=1), which assumes more than 14 symptomatic medications, or has a MIDAS score  $\geq 16$  points. During the pilot survey, when an interviewed clinician or patient encountered implausible combinations, they struggled with the choice tasks, and their responses were likely to be missed or unreliable. Thus, according to expert clinicians, some combinations were prohibited (see Online Supplementary Table S1). As is known, prohibitions hinder the achievement of a perfectly orthogonal and balanced design, but the applied statistical procedures allow for adjustments to such discrepancies from the optimal (but unrealistic) design. The tests of the DCE designs are presented in the Online Supplementary Materials (Tables S2 and S3).

### Sample size

According to Orme's formula (16), an appropriate sample size for DCE studies is  $n \geq 500 * b / (a * c)$  where  $a$  = the number of alternatives,  $b$  = the number of levels in the largest attribute, and  $c$  = the number of choice tasks. In our study we chose to compare just two profiles to facilitate the task (thus  $a = 2$ ),



**Figure 1.** Distribution of scores measured at T0 (before treatment) in each of the four considered measures included in the 4D migraine scale. To be noted that the distribution of MMDs is affected by rounding on 5-points scale (15, 20, 25, 30) and that MAMs and MIDAS appear normally distributed only after log-transformation.

**Table 1.** DCE design (attributes and levels).

Attribute	Representative chosen levels
Monthly Migraine Days (MMDs)	1 7 14 21 28
number of Monthly Acute Medications (MAMs)	1 7 14 21 28 35 42
Attack pain intensity Numerical Rating Score (NRS)	3 8 16 25 50 90
Migraine Disability Assessment (MIDAS)	1 3 5 7 9

the attribute with the largest number of levels was MAMs (thus  $b=7$ ) and the number of choice tasks was limited to 12 (thus,  $c=12$ ) to avoid incomplete compilations (drop outs), resulting in a sample size of at least 146. In addition, by using the Lighthouse Studio software simulator, we estimated the standard errors of the effects of each level obtainable with 300 respondents. We found that they were consistently below the recommended threshold of 0.05 (Online Supplementary Material, Table S3).

### Statistical analysis

Discrete choices of patients and clinicians were used to estimate the utilities that each respondent attributes to each attribute level. This estimation was performed according to the Hierarchical Bayes algorithm proposed by Lighthouse

Studio. Values above the reference line = 0 indicate a positive utility, whilst values less than 0 indicate a negative utility (disutility).

These measures were regressed vs. levels of each scale to identify the polynomial model that best interpolates the relationship between scores and utility. Only for the MIDAS, there was a necessity for a log-transformation to eliminate the lognormality of the original score distribution and to mitigate the potentially biasing effects of outliers. Since utilities were obtained within and among respondents, polynomial models were applied within the framework of General Estimating Equations, with the respondent serving as the “cluster”. The working correlation matrix was “unstructured”.

Through the estimation of the utilities associated with each possible value of each scale, an overall composite score, weighted by utility, was computed and standardized to a final scale from 0 (patient without any migraine) to 100 (patient with the worst possible migraine condition).

Since this computation was based on observed data, the formula will be reported and explained in the Results section.

This is not a clinical study since it does not deal with prevention/diagnosis/prognosis/therapy, and, according to Italian laws, it is not mandatory to obtain its approval by Territorial Ethical Committees.

Patients and physicians participating in the anonymized web-based survey provided their consent online. As the application of the 4D score for this retrospective efficacy

Which of these two patients do you think is in better condition?		
	Patient B	Patient A
Number of migraine medications in the past month	14	21
Disability due to migraine (MIDAS score: 0-5=minimal; 6-10=mild; 11-20=moderate; 21-39=severe; 40-79=very severe; >=80 extremely severe)	50	16
Pain intensity during worst attack in the past month (NRS: min=0, max=10)	5	7
Migraine days in the past month	21	7
My choice		
	<input type="radio"/>	<input type="radio"/>

**Figure 2.** Example of a choice task given to patients and clinicians who participated in the survey.

analysis on the previously collected data was considered in line with the objectives from the GARLIT Study (14) (approved by the Campus Bio-Medico University Ethical Committee n.30/20), the need to recall the patients to re-sign the informed consent form was waived.

## Results

### Estimation of relative importance and utility models

One hundred and ninety-seven migraine patients and 118 clinicians answered the questionnaire and completed the choice tasks. There was a substantial agreement between clinicians and patients about the weight of each parameter in terms of Relative Importance (RI) (Figure 3(a)). For both groups, MMD was the most important attribute (RI: 34% for clinicians, 32% for patients), while pain intensity-NRS was the least important one (RI: 14% for clinicians, 13% for patients). Though marginally, MIDAS was more important than MAMs for patients (29% vs. 26%), while for clinicians the relevance of these two attributes was almost equal (26% and 27%). In terms of the utility assigned to each level (Figure 3(b)), strong agreement was confirmed between clinicians and patients. Besides the trivial findings of decreasing utilities with increasing scores of each scale, some peculiarities deserve to be described: the highest utility is attributed to MMD=1, which is significantly larger than MAMs=1, MIDAS=3 and NRS=1; the highest disutility is attributed to MIDAS=90, even if only marginally higher (worse) than MMD=28 and MAMs=42. The pattern of the relationship between utility and scale levels is substantially linear only for MAMs: for MMD a steeper increase is observed moving from MMD=7 to MMD=1 with respect to changes between the other levels; MIDAS 3–8 are considered very close in terms of utility, as well as 16–25 and 50–90; NRS, which is confirmed the least

important attribute, shows similar utilities for the levels 3–5, with a slightly steeper increase from 3 and 1 and linear decrease from 5 to 9.

These patterns were further analyzed using polynomial models, which also allowed for the estimation of the utilities attributable to non-observed levels (interpolation between observed levels and extrapolation beyond observed levels). The equations were as follows:

$$Utility\_MMDs = 68.48 - 11.52 * \min(MMDs, 28) + 0.582 * \min(MMDs, 28)^2 - 0.011 * \min(MMDs, 28)^3 \quad (1.1)$$

$$Utility\_MAMs = 43.57 - 1.96 * \min(MAMs, 60) \quad (1.2)$$

$$Utility\_MIDAS = 51.43 - 4.836 * (\log_e(MIDAS + 1))^2 \quad (1.3)$$

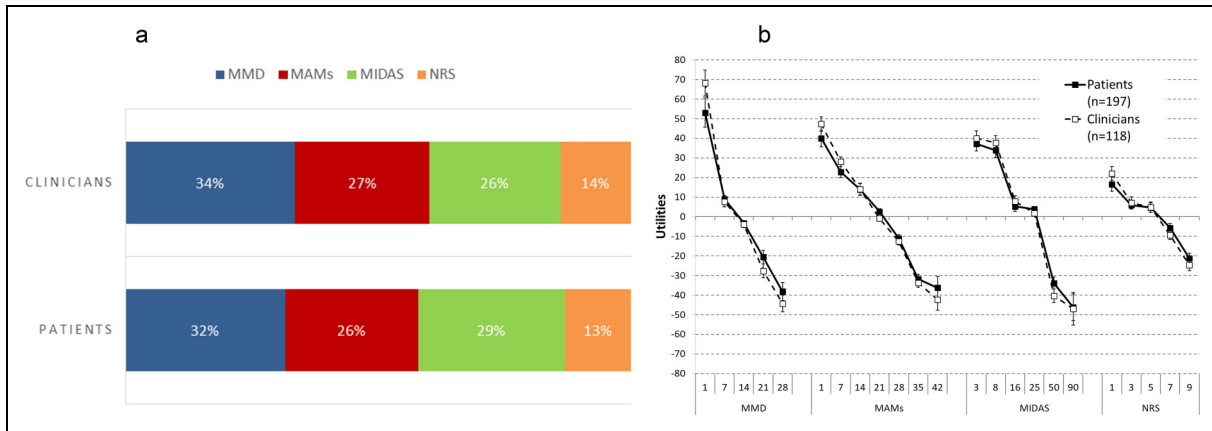
$$Utility\_NRS = 28.85 - 11.33 * NRS + 1.916 * NRS^2 - 0.144 * NRS^3 \quad (1.4)$$

The estimated relationships between utilities and scales are represented in Figure 4. According to the above equations, every possible value of each scale can be converted into a utility measure, and an overall utility can be derived by averaging the four utility measures.

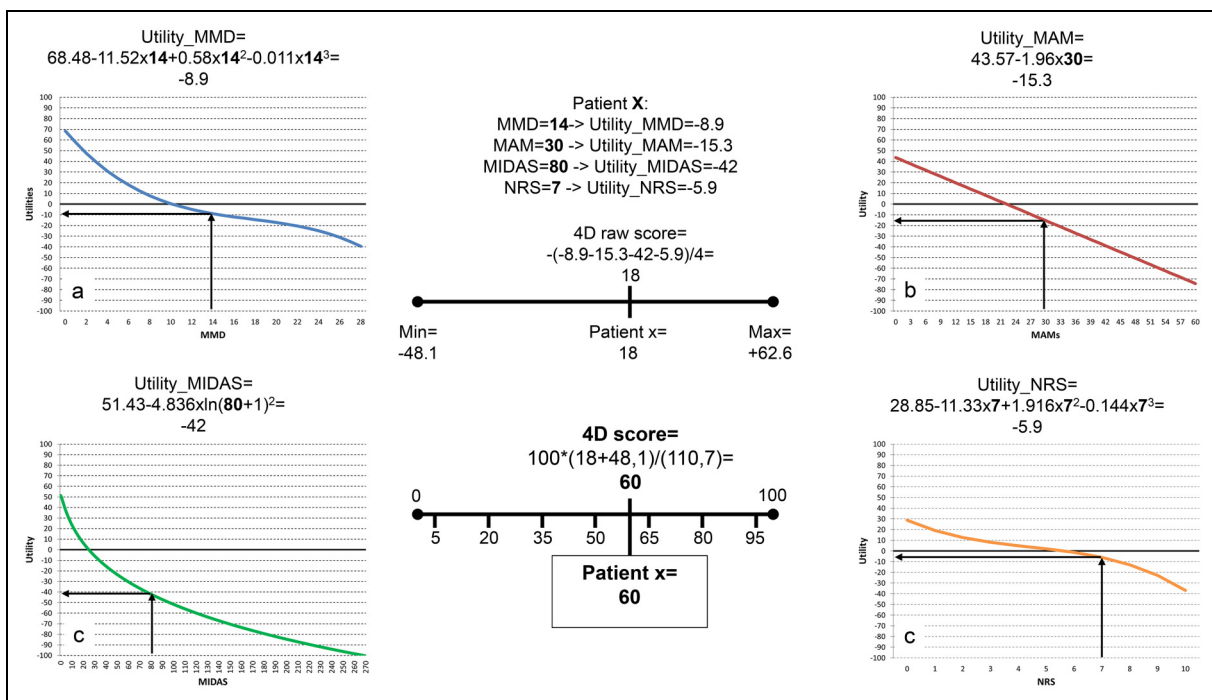
$$4D \text{ raw score} = -(Utility\_MMDs + Utility\_MAMs + Utility\_MIDAS + Utility\_NRS)/4 \quad (2)$$

To be noted the minus sign, which changes the direction of the scale, so the higher the 4D score, the worse the overall utility (as for each original measure).

In such a way, the overall utility can be interpreted as a weighted measure, where the weights are based on patients' preferences and confirmed by clinicians because of the strong agreement between patients and clinicians.



**Figure 3.** Weights of each parameter in terms of Relative Importance (RI) (a) indicating a good concordance between patients and clinicians; utilities assigned to each level (b): to be noted that, besides the concordance between patients and clinicians also in terms of level-specific utilities, the pattern is approximately linear only for MAMs, since for MMD the difference between 1 and 7 days is much higher than between the other consecutive pairs of levels, for MIDAS no relevant change of utilities was estimated between 3 and 8, between 16 and 25 and between 50 and 90, while utilities steeply decreased from 8 to 16 and from 25 to 50, for NRS again a simple linear relationship could not be assumed.



**Figure 4.** The estimated relationships between utilities and scales and an example of the 4D migraine composite score calculation for an individual patient X. Each panel shows the polynomial function used to convert one of the four migraine dimensions into its corresponding utility value, as estimated by the Discrete Choice Experiment (see Equations 1.1–1.4): (a) Utility of Monthly Migraine Days (MMDs); (b) Utility of Monthly Acute Medications (MAMs); (c) Utility of Migraine Disability Assessment (MIDAS); (d) Utility of Numerical Rating Scale (NRS) for pain intensity. For this exemplificative patient X, raw data are MMDs = 14, MAMs = 30, MIDAS = 80, NRS = 7, corresponding to utilities of –8.9, –15.3, –42, and –5.9, respectively. The mean of these utilities gives the 4D raw score of  $-(-8.9 - 15.3 - 42 - 5.9)/4 = 18$ . This value is then standardized to the 0–100 clinical range using Equation 3:  $4D \text{ score} = 100 \times (18 + 48.1)/(110.7) = 60$ . Horizontal scales represent the theoretical minimum (–48.1; corresponding to a migraine-free status) and maximum (+62.6; corresponding to the most severe condition). The resulting 4D score = 60 places the patient within the severe migraine category (50.1–65), according to the classification proposed (see Table 2).

**Table 2.** Computation of 4D score for three “extreme” patients (without migraine, with the mildest migraine possible, with the most severe migraine possible) and for a realistic “middle” patient X.

	patient without migraine	patient with mildest migraine	patient with worst migraine	Patient x
MMD	0	1	28	14
MAMs	0	0	60	30
MIDAS	0	0	270	80
NRS	0	1	10	7
utility MMD	68.5	57.5	-39.3	-8.9
utility MAMs	43.6	43.6	-74.2	-15.3
utility MIDAS	51.4	51.4	-100.3	-42.0
utility NRS	28.8	19.3	-36.8	-5.9
4D-score raw	-48.1	-43.0	62.7	18.0
4D-score (0–100)	0	5	100	60

Finally, to obtain a more intuitive, easy-to-use, and clinically relevant score, the following equation should be applied:

$$4D \text{ score} = 100 * (4D \text{ score raw} + 48.1)/110.7 \quad (3)$$

which allows for the representation of every possible patient on a 0–100 scale, ranging from 0 (a patient without migraine) to 100 (a patient with the most severe migraine possible) (Table 2 and Figure 4, which report an example of 4D migraine score calculation for an individual patient X).

#### 4D migraine score validation in a real-world setting

The 4D score was computed on a cohort of patients treated with galcanezumab for one year and evaluated every three months (14).

The 4D score is a preference-based average of four variables, and it was expected to tend towards a normal distribution (according to the central limit theorem, although, strictly speaking, in our case, the four components are not independent of each other). Since the established range was 0–100, to comprise 99.9% of patients (100% is not possible since the domain of the normal distribution is  $-\infty, +\infty$ ), a theory-based standard deviation is  $SD = (100-0)/(z_{0.9995}-z_{0.0005}) = 100/6.58 = 15.2$ . According to data gathered for galcanezumab, observation-based SDs resulted equal to 13.0 at T0 (baseline), 16.1 at T1 (after three months), 15.8 at T2 (after six months), 13.6 at T3 (after nine months) and 14.7 at T4 (after 12 months), with an average of 14.6 which confirms that  $SD=15$  could be assumed for 4D migraine score. With such SD we propose the following classes of 4D migraine scores: 0–5 “none”, 5.1–20 “very mild”, 20.1–35 “mild”, 35.1–50 “moderate”, 50.1–65 “severe”, 65.1–80 “very severe”, 81.1–95 “most severe”, 95.1–100 “extreme”.

Figure 5 represents the 4D migraine score distribution at each time of the galcanezumab real-world study, showing that baseline severe score values shifted toward mild-moderate classes three months after the introduction of galcanezumab and the benefit persisted throughout the one-year treatment.

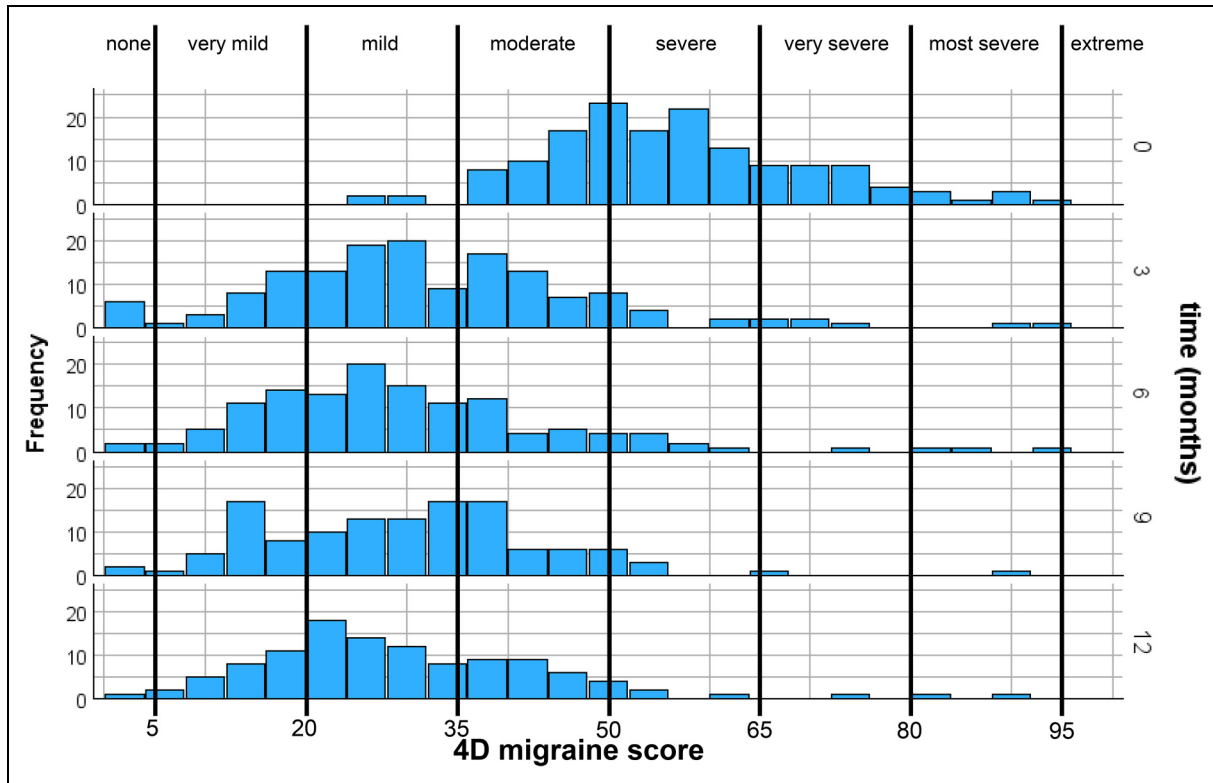
In addition, the estimate of the theoretical and empirically confirmed SD allows to lean on effect size conventions and to propose that  $0.8*15 = 12$  can be considered a “large” effect size (very relevant difference between two individuals as well as very relevant within-patient change),  $0.5*15 = 7.5$  a “medium” effect size and  $0.2*15 = 3$  a “small” effect size.

As shown in Figure 6, a strong effect of the treatment was found for each outcome (eta-squared consistently above 0.25 and  $p < 0.001$ ), and the composite 4D score properly summarized the changes after treatment with a larger effect size (eta-squared = 0.691) than that of each single parameter, suggesting higher sensitivity to changes. As a further check, the concurrent validation was assessed with respect to HIT-6, a patient-reported outcome measure not included in our 4D scale. As shown in Table 3, the concurrent validity of the 4D score is consistently higher than that of the single outcome measures, further confirming that the proposed composite score effectively summarizes and incorporates migraine patient status and outcome measures.

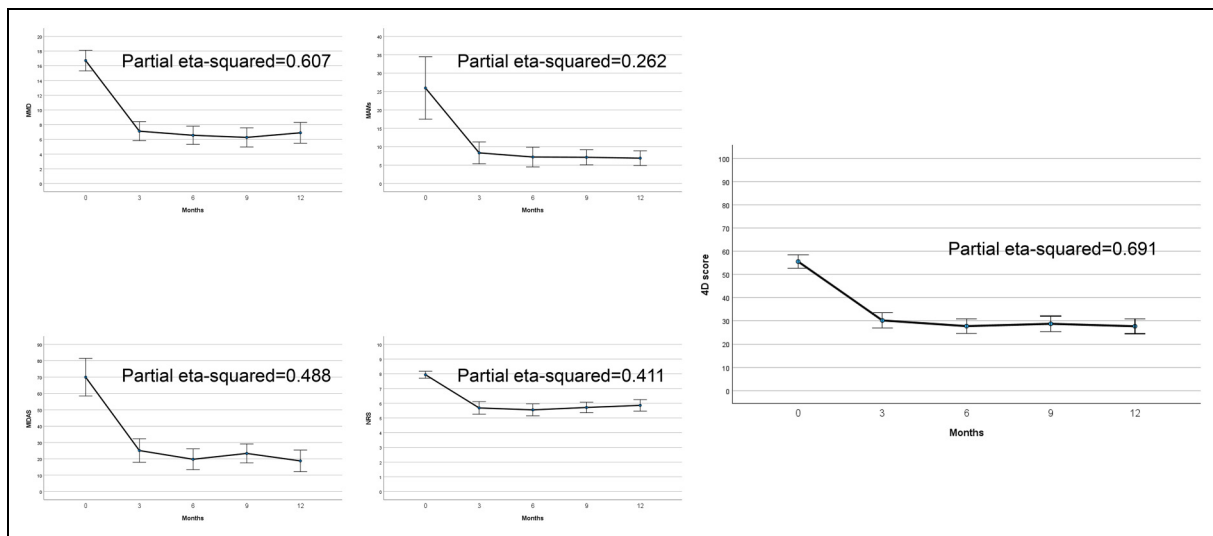
## Discussion

We aimed to create the 4D migraine composite scale, incorporating four of the most commonly used measures for evaluating the severity of migraine, MMDs, MAMs, pain intensity-NRS and MIDAS, and to evaluate its potential application in fully defining the migraine status and verifying the effectiveness of a preventive treatment. Our choice of the four dimensions was based on their widespread use and clinical relevance of these variables in both research and practice for assessing migraine burden and treatment efficacy. Furthermore, these variables are easy to collect, also allowing centers not dedicated to headache treatment to collect them and potentially use the 4D score.

Our scale contains two objective parameters, monthly migraine days, MMDs, and the number of monthly acute medications, MAMs, which are reliable in assessing the



**Figure 5.** Distribution of 4D migraine score at baseline and 3, 6, 9 and 12 months after the introduction of galcanezumab. Classes are defined according to standard deviation of 15, with the exception of the “none” and “extreme” classes.



**Figure 6.** Observed changes in the Real World study with galcanezumab: effect sizes of 4D composite score and of four single included parameters.

frequency of attacks and the patients’ need to treat them to restore their functionality to face daily activities. The parameter MMD was the most chosen by both patients and clinicians as the most important in determining migraine

burden. We included MAMs in the composite score because, in our opinion, the number of symptomatic drugs rather than the days with acute medications is more helpful in describing attack severity and response to acute

**Table 3.** Concurrent validity of 4D score

Correlation with HIT-6	4D score	MAMs	MMD	NRS	MIDAS
T0	<b>0.287</b>	0.161	0.179	0.254	0.221
3 months	<b>0.575</b>	0.393	0.494	0.485	0.496
6 months	<b>0.613</b>	0.522	0.586	0.577	0.399
9 months	0.471	0.422	<b>0.501</b>	0.362	0.416
12 months	<b>0.581</b>	0.506	0.548	0.49	0.502

Spearman's rank correlations between 4D score and HIT-6 at each follow-up evaluation (galcanezumab real-world data) compared with the correlations vs. single outcome scales.

Highest correlation for each time in bold.

treatment in terms of pain freedom. Then, we added two other measures: pain intensity and the MIDAS score. They are totally subjective from the patients' side as they describe the highest intensity level of migraine pain experienced by the patients and their headache-related disability in three life domains, work (or school-work), household work, and family, social, and leisure activities over a three-month time period. We chose to include these parameters, which are used worldwide in defining migraine severity, rather than creating different indicators or new multi-assessed scales, to establish a standardized common ground and terminology for evaluating migraine severity. Moreover, the relative weight of each parameter was rated based on both patients' and headache experts' opinions according to the DCE validated approach. The Discrete Choice Experiment is an accepted methodology for gathering and analyzing quantitative data on patients (17). It was applied in migraine field to understand preventive treatment preferences of patients with migraine (18,19). In our study, DCE was applied, and a definite score was attributed to each level of the included four parameters. The total 4D score was calculated by summing up the scores of each of the four parameters for each patient. To the best of our knowledge, this is the first time an outcome measure has been created that shares both patients' and clinicians' experiences.

Evaluating the clinicians' point of view too may have important clinical implications. As an example, some patients responding rapidly to acute medications, do not perceive an everyday disability and may underestimate their migraine burden. Conversely, the headache specialist may consider the frequent use of acute medications negatively due to the risk of developing medication overuse headache, which may influence therapeutic choices.

We did not want to eliminate or minimize the clinical value of the parameters considered singularly. Still, with the 4D scale we can evaluate the migraine severity from four different points of view, including parameters used worldwide in the research and clinical management of migraine, and combine them to obtain one score, i.e., one number, expressing a more complete quantification of patients' migraine severity than that explored by single parameters. The four components

are usually used as independent outcome measures in studies investigating the impact of treatments on migraine patients, and frequently they improved (or worsened) in parallel. However, they are often interrelated: MMDs are the basis for calculating the MIDAS score, and pain intensity is a subjective measure of disability itself; MAMs often reflect the days with headache but also the disability of patients' attacks, and so on. However, in our opinion, none of the parameters alone can fully evaluate migraine severity, and the relationship between them is only implicit. Considering the four of them in a composite scale allows for measuring migraine severity or the impact of a treatment comprehensively and also catching the interrelation among the parameters that each of them independently cannot detect. Moreover, it can assess the impact of migraine in patients without collinearity of such parameters (e.g., high MAMs and low MIDAS score). Finally, the composite 4D score reached an effect size of 0.691, which makes the change from baseline more visible and the effect of a treatment more consistent than each single measure.

Severity of migraine is an essential determinant of headache patients' disability, quality of life, and health care resource utilization. However, no validated method currently exists for globally measuring migraine burden. About 20 years ago, the MIGSEV scale attempted to identify items capable of assessing the severity of migraine with a high level of clinical and psychometric relevance and to combine these items into a unitary severity scale (20). The study consisted of a three-stage process. In the first step, the most relevant items for determining severity were defined a priori by consensus and built into a questionnaire. Secondly, the psychometric properties of the questionnaire were investigated in a prospective survey of migraine subjects, and the most discriminating severity items were identified by principal components analysis. Finally, the selected items were built into a unitary scale, and a consensus of neurologists devised a scoring system. In the study, seven items reflecting severity were identified: intensity of pain, tolerability, disability in daily activities, presence of nausea or vomiting, resistance to treatment, duration of attacks, and frequency of attacks. Principal components analysis identified three dimensions accounting for 65% of the overall variance, related to the intensity of attacks, resistance to treatment, and frequency of attacks. The internal consistency and the reproducibility of the items appeared quite satisfactory, and variations in these items were correlated with quality-of-life scores. The authors concluded that the items on the intensity dimension could be combined to generate a simple scale for assessing migraine attack severity. However, with the three dimensions identified, the MIGSEV scale measures the clinical severity of migraine attacks rather than their impact on patients' lives. Finally, no prospective study has been conducted to validate the MIGSEV scale, assessing its validity against patients' and physicians'

subjective assessments of severity and quantitative measures of quality of life.

Another more recent study (21) investigated how migraine patients can rate the severity of their disease using the Global Assessment of Migraine Severity (GAMS), that can be obtained by asking the patient to respond to a single question: “taking into account all aspects of your disease, how severe is your disease?” with seven response categories, from (1) not at all severe to (7) extremely severe, and how these ratings correlate with disability scores (MIDAS) or even with psychological comorbidities (in particular depression, measured using the 9-item Patient Health Questionnaire (PHQ-9) and the 14-item Hospital Anxiety and Depression Scale (HADS). However, the measure remained a single, subjective measure, from the patients’ side only, not combined with other parameters, and thus did not represent an objective evaluation of the migraine burden.

More recently, the concept emerged that it is the headache pain, in terms of duration and severity, rather than the headache frequency itself, that determines the level of headache disability. A ‘total pain burden (TPB)’ measure (12) including monthly headache days, duration of migraine headache, and maximum severity of migraine headache on a given day was proposed. A greater reduction in TPB was seen in patients with episodic or chronic migraine treated with galcanezumab relative to placebo. This enables a more comprehensive assessment of the potential benefits of preventive treatments in reducing the overall pain burden, thereby supporting patient-centered evaluations of treatment expectations.

Similarly, the ‘whole pain burden’ (13), derived from the product of headache frequency (i.e., mean of attacks per month in the last three months) per headache intensity (mean of NRS values) per headache duration (mean headache hours when treated) went in that direction. It was used, together with other PROMs on migraine-related disability, impact, and quality of life, as well as symptoms of depression or anxiety, to evaluate the galcanezumab impact on patients who had experienced previous unsuccessful preventive treatments.

The proposals above remarked that a composite measure may define migraine severity but also evaluate the effectiveness of migraine treatments based on more than one parameter. However, both proposed composite scales focused mainly on the attack characteristics, and they did not include a disability measure in their evaluation.

Moreover, an isolated parameter may not fully reflect the clinical improvement in its complexity after treatment, as it may also depend on several other features. In other words, the best clinical approach for physicians when treating a patient should not be based solely on a single parameter in the decision to prolong or interrupt that therapy. This is well demonstrated by a paper from De Icco et al. (22). According to the Italian Medicines Agency (AIFA) rules, anti-CGRP monoclonal antibodies and gepants in Italy are prescribed for the preventive treatment of migraine in

patients who have the following criteria: i) at least eight migraine days per month in the last three months, ii) MIDAS score  $\geq 11$ , and iii) previous failure due to lack of efficacy or tolerability of at least three standard of care oral preventive drugs. Although recently updated, the treatment could be continued only in patients who reported  $a \geq 50\%$  reduction in MIDAS score after an initial period of three months. Thus, in Italy, the MIDAS score represents one of the key indicators for treatment initiation and a limiting factor for treatment continuation. The authors evaluated the performance of MIDAS and MMDs in achieving  $a \geq 50\%$  reduction after three months of treatment in predicting long-term (1-year) clinical effectiveness. The pivotal finding of the study was that up to one-third of patients qualifying as responders at the end of one-year treatment would have been discontinued early if the  $\geq 50\%$  MIDAS reduction criterion alone had been applied. The same would also be the case if the unique MMD parameter were considered. By contrast, the combination of these two parameters provided an improved accuracy: if the “either MIDAS or MMDs” criterion were applied, only 16% of responders at T12 would have been discontinued after only three months of erenumab treatment. Once more, this study confirms that ideally the combination of several features, namely: reduction in headache frequency, pain intensity levels, pain duration, acute drugs consumption, as well as improvement in patients’ reported outcomes such as disability and quality of life may fully describe both the migraine burden and the clinical response to a treatment in its complexity, including the option to continue or discontinue that treatment. This is particularly true in hard-to-treat migraine patients, where a mild reduction in all four composite scale parameters results in clinically meaningful improvement (23).

The 4D score could be adapted to different scenarios. In the present study, we used it to track meaningful clinical changes over time during one-year treatment with galcanezumab. The score allows longitudinal monitoring of clinically relevant change through a single composite measure. However, the 4D score could also be useful for clinical decision-making at baseline, by contextualizing migraine burden when individual parameters are discordant, and potentially supporting decisions such as the initiation of preventive therapy. In this context, the score serves as an additional standardized metric to aid decision-making, complementing existing guideline-based criteria.

The study has some limitations. The estimate was obtained with statistical models that accounted for the preferences of a limited sample of about 300 people with migraine and headache experts. Moreover, we had to rely on model-derived utilities to define the score based on the weights of patients and clinicians. A discrete choice experiment is a robust and validated method to assess preferences and to transform choices into utilities. However, the method has some limitations, the most important are: 1) it is based

on self-completed questionnaires and requires that participants carry out several choice tasks, with the consequence that motivated participants provide evaluable answers with the risk of a selection bias; 2) preferences and utilities strongly depend on how attributes and levels are proposed, i.e., the inclusion/exclusion of an attribute or a different choice of levels within an attribute could modify the estimation of relative importance and utilities. Concerning limitation 1), we could not exclude a selection bias of highly motivated patients but, thanks to the administration of the questionnaire to clinicians also, we documented a close similarity of preferences in the two groups of participants, which gives us the possibility to venture a certain generalizability of our results, even if it needs to be verified in an independent sample. Considering limitation 2), the four components (MMDs, MAMs, MIDAS, and Pain-NRS) do not cover all possible domains of migraine, but they are the most considered in RCTs and Real-World studies. In addition, the levels were carefully chosen to represent the

whole potential range of each variable; thus, the bias associated with the choice of specific and arbitrary levels should not be relevant in our study.

Finally, 4D is still a concise score, which could be a weak point but also a strong point, given its higher sensitivity compared to any single parameter.

## Conclusions

The 4D migraine scale, based on the preference weights of clinicians and patients, could be a useful outcome measure to be considered in clinical trials for fully measuring the migraine burden and the efficacy of preventive treatments. Further studies on larger samples of migraine patients are needed to verify the viability of such a composite score in clinical practice for treatment indication and monitoring in terms of clinical outcome, patients' satisfaction, impact on the interictal burden, and social costs.

### Clinical implications

- A single measure may only partially represent the burden of migraine.
- We created the composite 4D migraine scale, including four of the most used parameters: monthly migraine days, monthly acute medications, pain intensity, and Migraine Disability Assessment Score.
- The 4D scale was based on the preference weights of both clinicians and patients. It resulted in a more sensitive score than single parameters to quantify migraine severity and treatment efficacy fully.

### Acknowledgments

The authors want to thank Alleanza Cefalalgici (Al.Ce.), an Italian Migraine Patients Association, all patients, and all headache experts who answered the DCE questionnaires.

### Author contributions

Study concept and design: FV and PP. Acquisition of data: FV, CA, LF, LFI, MM, GB, MR, SC. Analysis and interpretation of data: FV, GCM, and PP. Drafting of the manuscript: FV and PP. Revising it for intellectual content: FV, GCM, CA, LFI, MR, PP. Final approval of the completed manuscript: FV, LFI, MR, SC, CA, GCM, and PP.

### Consent to participate

Patients and physicians participating in the anonymized web-based survey provided their consent online. As the application of the 4D score for this retrospective efficacy analysis on the previously collected data was considered in line with the objectives from the GARLIT Study (approved by the Campus Bio-Medico University Ethical Committee n.30/20), the need to recall the patients to re-sign the informed consent form was waived.

### Data availability statement

Data supporting the findings in the present study are reported in the article and the supplementary materials. The data collected

and analyzed for the current study are available from the corresponding author on reasonable request.

### Declaration of conflicting interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: CA received honoraria for scientific lectures and presentations from AbbVie, Pfizer, and Teva; consulting fees from AbbVie, Lundbeck, and Pfizer; he is Principal Investigator of a funded research project from Dysport and a funded clinical trial from Teva. LF received honoraria for scientific presentations and travel fees from AbbVie, Eli Lilly, Lundbeck, Novartis, Organon, Pfizer, and TEVA. LFI received financial support, honoraria for scientific lectures and presentations, consulting fees for the participation in advisory boards, and support for attending meetings from: Teva, Eli Lilly, Lundbeck, Pfizer, Organon, and AbbVie; he is Associate Editor for *Frontiers in Neurology*, Headache and Neurogenic Pain section and Review Editor for *Therapeutic Advances in Neurological Disorders*. MR has received research support, speaker honoraria, and support to attend national and international conferences from Lundbeck, Pfizer, AbbVie, Jazz Pharmaceuticals, and Angelini Pharma. SC received honoraria for speaker panels from Teva, AbbVie, Organon, Pfizer, and Lundbeck. FV has received financial support from AbbVie, Angelini and Lundbeck for investigator-initiated trials; consulting fees for the participation in advisory boards from AbbVie, Angelini, Eli Lilly, Lundbeck, Organon,

Novartis, Pfizer, and Teva; honoraria for scientific lectures and presentations from AbbVie, Eli Lilly, Lundbeck, Novartis, Pfizer, and Teva; support for attending meetings from Abbvie, Eli Lilly, Lundbeck, Pfizer, and Teva; has been Principal Investigator in clinical trials sponsored by AbbVie, Eli Lilly, Lundbeck, Pfizer, and Teva; he is co-Specialty Chief Editor for *Frontiers in Neurology*, Headache and Neurogenic Pain section.

PP, MM, GB, and GM reported no conflict.


### Ethical considerations

This study was approved and granted by Sapienza University of Rome (Progetti di Ateneo/University projects, call 2021, Protocol: RP12117A8A841167).

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study had a grant from Sapienza University of Rome (Progetti di Ateneo/University Project, call 2021, Protocol: RP12117A8A841167).


### ORCID iDs


Patrizio Pasqualetti  <https://orcid.org/0000-0001-5560-1979>


Claudia Altamura  <https://orcid.org/0000-0002-5934-5535>

Luisa Fofi  <https://orcid.org/0000-0001-7958-5440>

Luigi Francesco Iannone  <https://orcid.org/0000-0003-3598-0195>

Marilena Marcosano  <https://orcid.org/0009-0003-5191-7628>

Marina Romozzi  <https://orcid.org/0000-0002-7071-1321>

Fabrizio Vernieri  <https://orcid.org/0000-0002-9594-9336>

### Supplemental material

Supplemental material for this article is available online.

### References

- Lipton RB and Silberstein SD. Episodic and chronic migraine headache: breaking down barriers to optimal treatment and prevention. *Headache* 2015; 55: 103–122.
- Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. *Cephalalgia* 2018; 38: 1–211.
- Ishii R, Schwedt TJ, Dumkrieger G, et al. Chronic versus episodic migraine: the 15-day threshold does not adequately reflect substantial differences in disability across the full spectrum of headache frequency. *Headache* 2021; 61: 992–1003.
- Manzoni GC, Bonavita V, Bussone G, et al. Chronic migraine classification: current knowledge and future perspectives. *J Headache Pain* 2011; 12: 585–592.
- Ashina M. Migraine. *N Engl J Med* 2020; 383: 1866–1876.
- Sacco S, Amin FM, Ashina M, et al. European headache federation guideline on the use of monoclonal antibodies targeting the calcitonin gene related peptide pathway for migraine prevention – 2022 update. *J Headache Pain* 2022; 23: 67.
- Ferreira-Valente MA, Pais-Ribeiro JL and Jensen MP. Validity of four pain intensity rating scales. *Pain* 2011; 152: 2399–2404.
- Alpuente A, Torres-Ferrus M, Caronna E, et al. The state of art on the use of patient reported outcomes in migraine. *Curr Opin Neurol* 2024; 37: 271–282.
- Stewart WF, Lipton RB, Kolodner K, et al. Validity of the migraine disability assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain* 2000; 88: 41–52.
- Kosinski M, Bayliss MS, Bjorner JB, et al. A six-item short-form survey for measuring headache impact: the HIT-6. *Qual Life Res.* 2003; 12: 963–974.
- Rendas-Baum R, Bloudek LM, Maglinte GA, et al. The psychometric properties of the migraine-specific quality of life questionnaire version 2.1 (MSQ) in chronic migraine patients. *Qual Life Res* 2013; 22: 1123–1133.
- Ailani J, Andrews JS, Rettiganti M, et al. Impact of galcanezumab on total pain burden: findings from phase 3 randomized, double-blind, placebo-controlled studies in patients with episodic or chronic migraine (EVOLVE-1, EVOLVE-2, and REGAIN trials). *J Headache Pain* 2020; 21: 123.
- Silvestro M, Tessitore A, Orologio I, et al. Galcanezumab effect on “whole pain burden” and multidimensional outcomes in migraine patients with previous unsuccessful treatments: a real-world experience. *J Headache Pain* 2022; 23: 69.
- Vernieri F, Brunelli N, Marcosano M, et al. GARLIT Study Group. Maintenance of response and predictive factors of 1-year Galcanezumab treatment in real-life migraine patients in Italy: the multicenter prospective cohort GARLIT study. *Eur J Neurol* 2023; 30: 224–234.
- Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health* 2011; 14: 403–413.
- de Bekker-Grob EW, Donkers B, Jonker MF, et al. Sample size requirements for discrete-choice experiments in health-care: a practical guide. *Patient* 2015; 8: 373–384.
- Hauber AB, González JM, Groothuis-Oudshoorn CG, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value Health* 2016; 19: 300–315.
- Mansfield C, Gebben DJ, Sutphin J, et al. Patient preferences for preventive migraine treatments: a discrete-choice experiment. *Headache* 2019; 59: 715–726.
- Schwedt TJ, Martin A, Kymes S, et al. Patient preferences for attributes of injected or infused preventive migraine medications: findings from a discrete choice experiment. *Headache* 2023; 63: 484–493.
- El Hasnaoui A, Vray M, Richard A, et al. MIGSEV Group. Assessing the severity of migraine: development of the MIGSEV scale. *Headache.* 2003; 43: 628–635.
- Sajobi TT, Amoozegar F, Wang M, et al. Global assessment of migraine severity measure: preliminary evidence of construct validity. *BMC Neurol* 2019; 19: 53.
- De Icco R, Vaghi G, Allena M, et al. Does MIDAS reduction at 3 months predict the outcome of erenumab treatment? A real-world, open-label trial. *J Headache Pain* 2022; 23: 123.
- Altamura C, Cevoli S, Brunelli N, et al. GARLIT Study Group collaborators. When should we consider chronic patients as non-responders to monoclonal antibodies targeting the CGRP pathway? *J Neurol* 2022; 269: 1032–1034.